Proc. Natl. Acad. Sci. USA
Vol. 78, No. 4, pp. 2150–2154, April 1981
Biochemistry

Secondary structure of eukaryotic cytoplasmic 5S ribosomal RNA

(coaxial helices/*Chlorella* 5S rRNA/*Scenedesmus* 5S rRNA)

KENNETH R. LUEHRSEN AND GEORGE E. FOX

Department of Biophysical Sciences, University of Houston, Houston, Texas 77004

Communicated by James E. Darnell, Jr., December 24, 1980

ABSTRACT **A five-helix secondary structural model is proposed for eukaryotic cytoplasmic 5S rRNA. All available sequence data are consistent with this model including those from *Chlorella* 5S rRNA whose sequence is revised by data included here. Various architectural features of eukaryotic 5S rRNA are summarized in terms of this secondary structural model. It is observed that previous failures to identify universal models for 5S rRNA secondary structure stem from significant differences in architecture between eukaryotic cytoplasmic and eubacterial 5S rRNAs. The usual four-helix model for eubacterial 5S rRNA secondary structure nevertheless does share several structural features with the five-helix model presented here for cytoplasmic 5S rRNA. It is thus likely that these two classes of 5S rRNA are the result of evolutionary divergence rather than convergence.**

The secondary structure of 5S rRNA has been the subject of extended discussion (1). Nevertheless, general agreement has remained elusive. In the case of prokaryotes, comparison of published sequences continues to support the four-helix model of Fox and Woese (2, 3), and that model with minor modifications (4) is now widely utilized. Of these four helical regions, three are found to be universal among eukaryotic sequences as well. The fourth structure, a stable hairpin with a short (usually three or four nucleotides) single-stranded loop is found only among bacterial 5S rRNAs and has thus been referred to as the prokaryotic loop. The presence of a structural feature that is unique to prokaryotic 5S rRNA makes it clear that, although prokaryotic and eukaryotic 5S rRNAs share common ancestry, they are structurally distinct. This view is supported by biochemical studies that demonstrate that eukaryotic and prokaryotic 5S rRNAs are not functionally interchangeable (5, 6).

A real remaining issue is to determine what, if any, additional secondary structural features exist in eukaryotic 5S rRNA in lieu of the characteristic prokaryotic hairpin loop. Several alternatives have been suggested. One (7, 8) involves pairing between positions 78–82 and 95–99 of human KB cell 5S rRNA. The other (8, 9) would pair positions 66–72 with 103–109. Herein we argue that, in fact, the secondary structure of eukaryotic cytoplasmic 5S rRNA is properly represented by a five-helix model which includes both of these helical regions as well as the three that have been established previously. This five-helix secondary structural model is applicable to all known eukaryotic 5S rRNA sequences except the sequence of *Chlorella*.

In this communication we report experimental evidence that necessitates revision of the *Chlorella* 5S rRNA sequence. When so corrected, it also conforms to the proposed secondary structure model.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

5S rRNA was isolated and purified from *Scenedesmus quadricauda* (ATCC 11460) and *Chlorella* sp. (ATCC 11469) as described (10). Both 5′ and 3′ *in vitro* ^{32}P-end-labeled RNAs were obtained in the usual manner (11, 12). Sequence analysis was conducted from both termini by the enzymatic method (13) and from the 3′ terminus by the chemical method (14). Distinction between pyrimidines on the enzymatic gels was greatly facilitated by the use of the recently described *Physarum* M nuclease activity (15). Resolution near the 5′ terminus was improved by the addition of a four-nucleotide extending sequence.

RESULTS

Sufficient gel ladders were run to determine unequivocally the sequence of *Chlorella* 5S rRNA in its entirety. This revised sequence is reported in Fig. 1. A second series of sequence analysis gels run on *Scenedesmus* 5S rRNA established that its sequence is identical to that of *Chlorella* with the likely exception of position 93 which may be a cytosine in *Scenedesmus* (only an enzymatic gel was available in this region and these are not al-

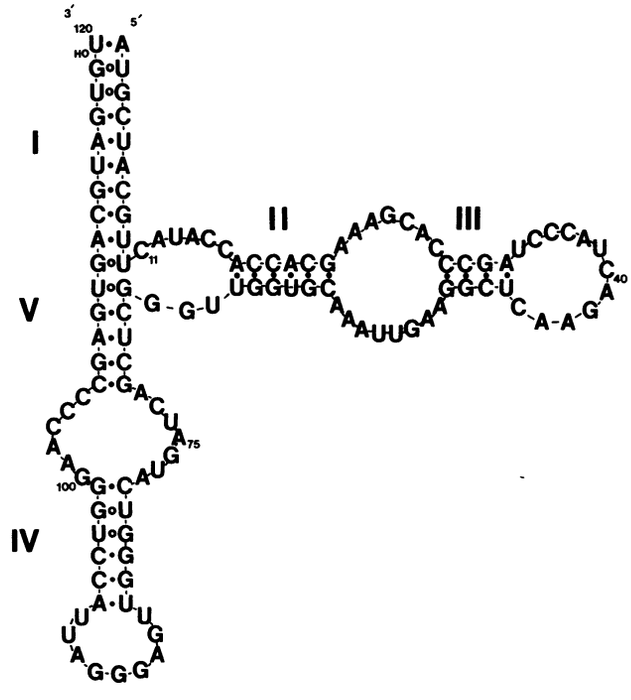

FIG. 1. The sequence of *Chlorella* 5S rRNA, as revised by data presented herein, drawn in the five-helix secondary structure described in the text.

ways reliable in making the cytosine/uracil distinction). The *Chlorella* sequence as determined here differs from the earlier published version (16, 17) and should properly be regarded as correcting that work. In detail the changes are as follows. (*i*) The

$$\overset{107}{\text{-G-A-G-U-}}$$

sequence segment is properly located immediately after

$$\overset{98}{\text{-G-G-G-A-A-C-C-C-C-}}$$

instead of before it. (*ii*) The pancreatic RNase oligomer

$$\overset{86}{\text{-G-A-G-G-G-A-U-}}$$

was previously reported as -G-G-A-G-G-A-U-. (*iii*) The RNase T1 oligomer

$$\overset{9}{\text{-U-U-C-A-U-A-C-C-A-C-C-A-C-G-}}$$

was previously reported as -U-U-C-A-U-A-C-A-C-C-A-C-G-. Sequence analysis gels documenting these corrections are shown in Figs. 2 and 3.

DISCUSSION

Eukaryotic 5S rRNA Secondary Structure and Its Implications for Tertiary Structure. The corrections to the *Chlorella* 5S rRNA sequence reported here and the recently completed *Tetrahymena thermophila* sequence (19) necessitated a reexamination of the comparative evidence pertaining to the secondary structure of eukaryotic 5S rRNA. When this was done, five helical regions were identified (Fig. 1) that meet the comparative criteria. That is, (*i*) they can be formed in all the available sequences at comparable locations and (*ii*) the pattern of base pairing is conserved while the relevant sequence regions exhibit considerable phylogenetic variation. Indeed, of the five helical regions identified, only helix IV of Fig. 1 can be considered to be under any doubt. Here there is less variation between sequences than one might desire, and formation of the helix requires as many as three contiguous G·U pairs in the higher eukaryotes unless a U·U mispair is included. Sequences from the lower eukaryotes yeast and *Tetrahymena*, however, do show substantial variation in this region and the proposed helix is conserved under these conditions.

Fig. 4 shows various published eukaryotic 5S rRNAs (21) in the proposed secondary structural arrangement. Recently revised sequences are utilized for *Drosophila* (22) and the higher plant (18). In the case of yeast, a few sites of susceptibility to chemical modification have been identified (20) and these are consistent with the proposed structure. The recently determined sequence of sea urchin 5S rDNA (23) is also consistent with the five-helix structure.

The secondary structural model in Figs. 1 and 4 is drawn in such a way that helices I and V are coaxial. Precedent for such coaxial stacks exist in the known tertiary structure of tRNA. It would not be unreasonable to anticipate that such coaxial stacking ultimately may be found in eukaryotic cytoplasmic 5S rRNA as well.

Possible Extension of Helix II. It has been drawn to our attention (24) that, if one considers the cytosine at position 63 in the human sequence to be looped out, helix II can be extended to eight base pairs. Such an extension has been proposed previously (8) and is in fact supported by comparative evidence in most instances (there are exceptions among prokaryotic 5S rRNA sequences). Should this extension be included in the secondary structure? If the comparative evidence continues to be favorable, possibly it should be. Of course, it can be argued that the extension is best regarded as a tertiary structural feature. This would certainly be true if the two additional pairs were ultimately found not to be coaxial with helix II or if they were not of the Watson–Crick type.

It should be realized that this unresolved detail may be of considerable importance. At first sight it might seem that the extension involves little more than an argument about the length of helix II in Fig. 2. This is not so. With the extension included, all available eukaryotic 5S rRNA sequences can also be drawn in an alternative secondary structure format in which helices II and V are coaxial instead of I and V. Hence, if the extension is real, one must decide which secondary structural format is in fact most consistent with the actual tertiary structure. Not ruled out either is that a switch between two such stacking arrangements may be of functional significance.

Architectural Characteristics of Eukaryotic 5S rRNA. The availability of a general model for the secondary structure of eukaryotic cytoplasmic 5S rRNA allows inquiry into other aspects of structure. Fig. 5 schematically locates conserved and semiconserved nucleotides based on currently available sequence data. Although this listing will be subject to revision as more sequences are determined, it is apparent that the extent of conservation is much greater in the single-stranded regions. The structural domain defined by helices II and III is likewise more conserved than that defined by helices IV and V.

As in tRNA, the lengths of the various single-stranded and paired regions are quite characteristic. Table 1 summarizes these relative to the nomenclature of Fig. 5. Detailed comments and exceptions will be noted here. The overall length of the coaxial stalks of helices I and V is usually 15 or 16 pairs, and there is always a G·U pair at the point of abutment. The conserved hairpin loop, single-stranded region C, has 12 nucleotides, which is characteristically shorter than the eubacterial equivalent which typically has 13. Region E-1 contains the conserved sequence -A-G-U-A-. The -U-A- portion of this segment is universally found among eubacterial 5S rRNAs as well. Region E-2 contains two conserved adenines which are preceded by a guanosine in all published sequences. The guanosine is either part of helix IV or in region E-2. More constancy in the characteristic lengths of regions E-2 and D is obtained in the higher eukaryotes if a U·U mispair is allowed in helical region IV. For example, in human KB this is accomplished by pairing positions 94–98 (instead of 95–99) with positions 78–82.

Occasional mispairs are observed in some of the helical regions. For example, region II contains a G·G in trout and an A·C in *Crithidia* (25). In the case of yeast, region III can only be formed if one allows a looped-out nucleotide. Besides restoring the pairing, this looped-out base also restores single-stranded regions B_1, B_2, and C to their characteristic lengths. Occasionally, too, a helical region would appear to be one base shorter or longer than is usual. When this occurs the lengths of the neighboring single-stranded regions are also affected. An example is region II in *Tetrahymena*, which can be extended to seven pairs. If this actually occurs, then regions A-1 and A-2 each would be shortened by one nucleotide.

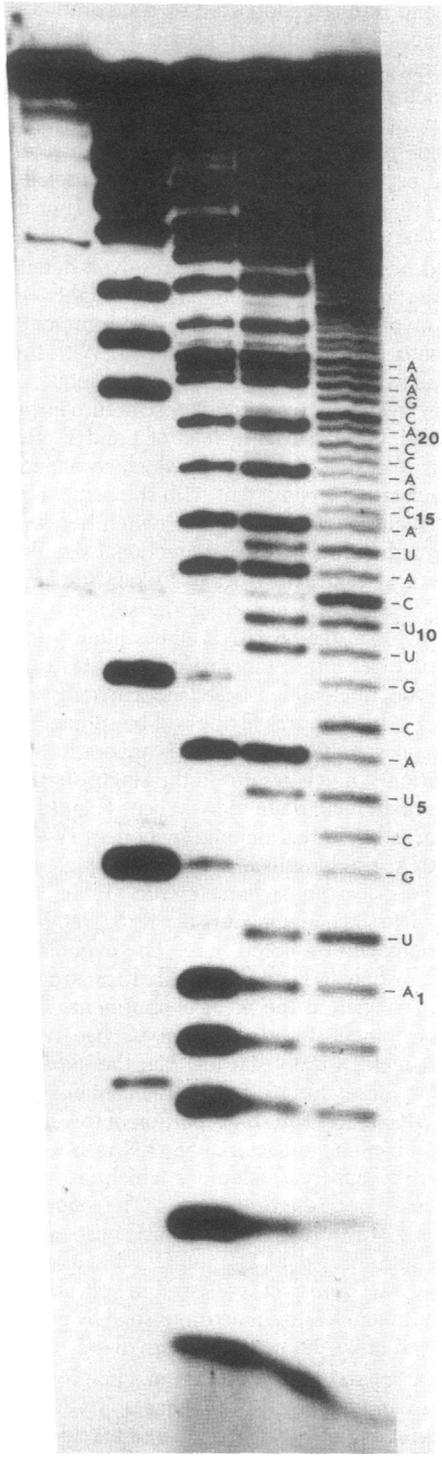

FIG. 2. Enzymatic sequencing ladder from *Chlorella* 5S rRNA. Prior to 5′-end-labeling, the 5S rRNA was extended with an oligoriboadenylate, $(Ap)_3A$, through the use of T4 RNA ligase. This facilitates the reading of the first few nucleotides. The gel itself was 20% polyacrylamide (40 cm × 0.45 mm). From left to right the lanes are: control, RNase T1, RNase U2, *Physarum* RNase M, and alkaline hydrolysis. This gel documents the sequence revision in the large oligomer, -U-U-C-A-U-A-C-C-A-C-C-A-C-G-.

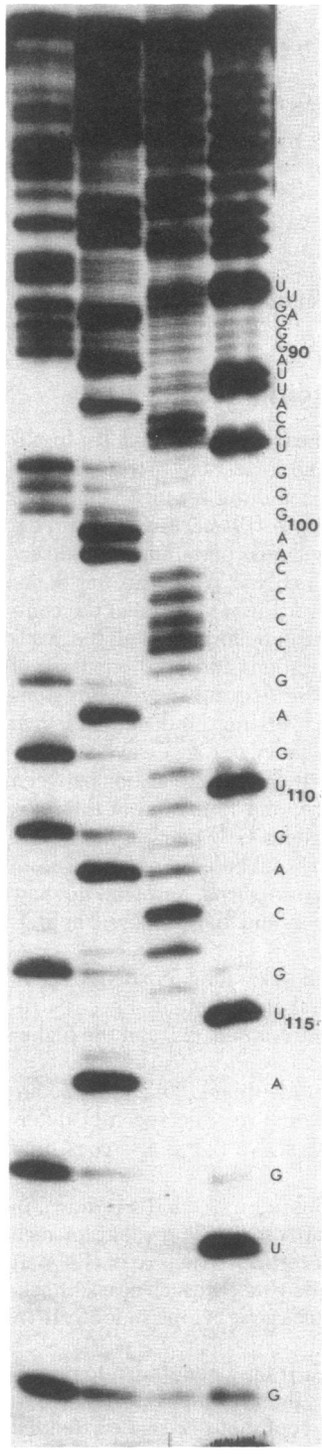

FIG. 3. Sequencing ladder for *Chlorella* 5S rRNA determined on 3′-end-labeled material by the chemical method. The gel was 20% polyacrylamide (40 cm × 0.45 mm). From left to right the lanes are: dimethyl sulfate (guanosine), diethylpyrocarbonate (adenosine > guanosine), and anhydrous hydrazine in 3 M NaCl (cytidine > uridine) and hydrazine (uridine). This gel documents the correct placement of -G-A-G-U- and the correct sequence of the pancreatic RNase oligomer -G-A-G-G-G-A-U-. The chemical method is known to produce artifact bands in the cytidine > uracil lane (18) and the present example is no exception. This tendency to read extra cytidine makes it desirable to verify the lengths of cytidine stretches with the enzymatic method. This has been done here (gels not shown) and the results obtained unequivocally support the interpretation shown in this figure.

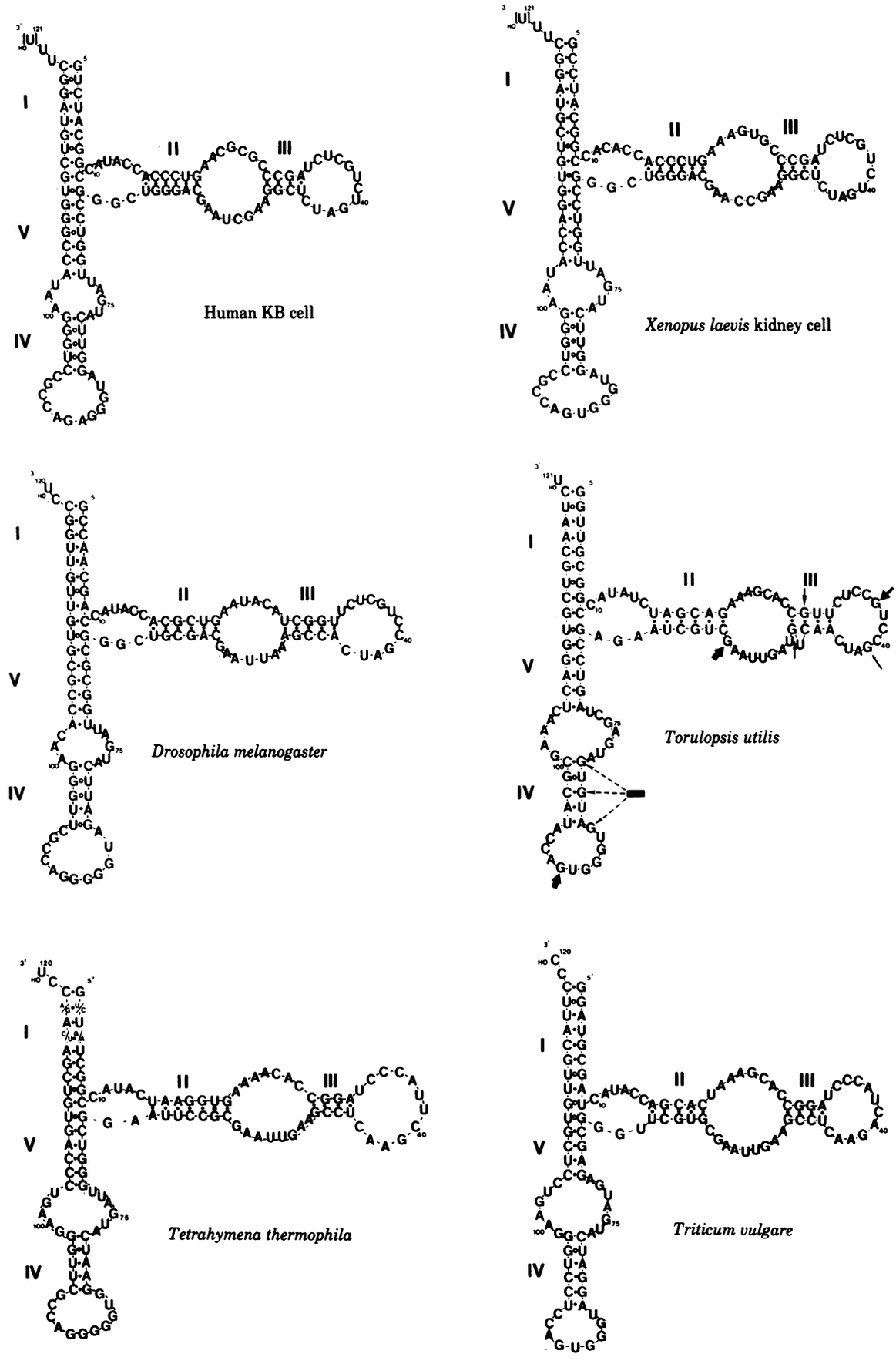

FIG. 4. Six diverse eukaryotic 5S rRNA sequences drawn in the proposed secondary structural format. Arrows are included on the yeast figure to indicate points of susceptibility to chemical modification (20). The *Triticum vulgare* sequence was determined by MacKay *et al.* (18) and serves to correct errors in earlier work on the higher plant 5S rRNAs.

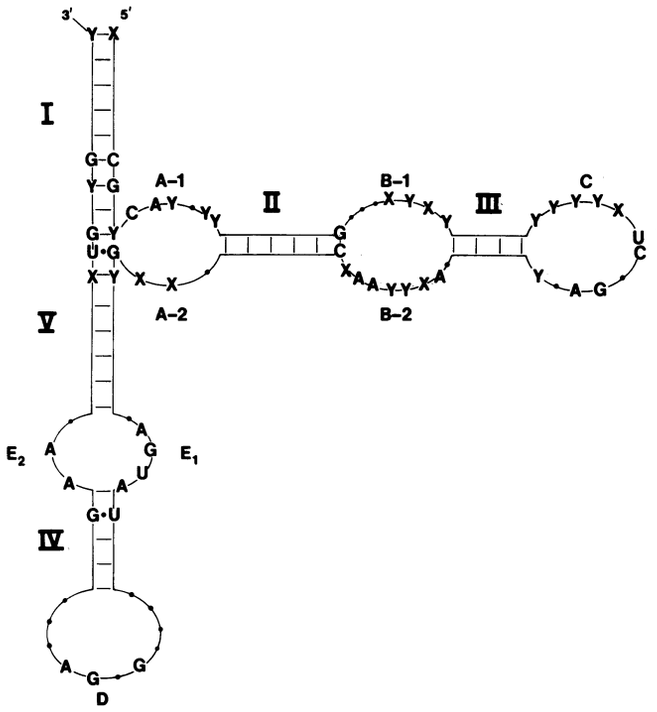

FIG. 5. Diagram indicating sites of conserved and semiconserved nucleotides in eukaryotic 5S rRNA. Single-stranded segments are designated as indicated in Table 1 (i.e., A-1, A-2, B-1, B-2, C, D, E-1, and E-2). Because no universal numbering system is available for eukaryotic 5S rRNA, dots are used to relate the numbers of bases in variable segments to the human KB cell sequence. Semiconserved bases are represented by X for purines and Y for pyrimidines.

Table 1. Usual architectural features of eukaryotic cytoplasmic 5S rRNA

Helical region	No. of pairs	Single-stranded region	No. of bases
I	9–10	A-1	6
II	6	A-2	3
III	4	B-1	7
IV	5–6	B-2	8
V	5–7	C	12
		D	9–12
		E-1	5–7
		E-2	3–6

Note Added in Proof. The five-helix model presented here has been independently found applicable to individual sequences by several investigators (A. G. Hinnebusch, L. C. Klotz, R. L. Blanken, and A. R. Loeblich III, personal communication; refs. 19, 24, and 26). What is uniquely documented herein is that this model is in fact generally applicable to all known eukaryotic sequences when various sequence errors are corrected. It should also be realized that, like the earlier eubacterial model (2, 3), this five-helix structure is properly regarded as a minimal model. It does not attempt to account for tertiary interactions and may not even include all interactions that are properly regarded as part of the secondary structure. Especially noteworthy in this regard are the extensions of helices II and IV that have been proposed (24). Such short segments are extremely difficult to evaluate by comparative data alone.

We are especially indebted to W. F. Doolittle, R. M. MacKay, D. F. Spencer, and M. W. Gray for providing us access to sequence data from wheat embryo and *Crithidia* 5S rRNA prior to publication. We are also grateful to H. Donis-Keller and N. R. Pace for their generous gifts of *Physarum* M nuclease and polynucleotide kinase. This work was supported by grants from the National Aeronautics and Space Administration (NSG-7440) and The Robert A. Welch Foundation (E-784) to G.E.F.

1. Erdmann, V. A. (1976) *Prog. Nucleic Acid Res. Mol. Biol.* **18**, 45–90.
2. Fox, G. E. & Woese, C. R. (1975) *Nature (London)* **256**, 505–507.
3. Fox, G. E. & Woese, C. R. (1975) *J. Mol. Evol.* **6**, 61–76.
4. Noller, H. F. & Garrett, R. A. (1979) *J. Mol. Biol.* **132**, 621–636.
5. Wrede, P. & Erdmann, V. A. (1973) *FEBS Lett.* **33**, 315–319.
6. Bellemare, G., Vigne, R. & Jordan, B. (1973) *Biochimie* **55**, 29–35.
7. Hori, H. (1976) *Mol. Gen. Genet.* **145**, 119–123.
8. Nishikawa, K. & Takemura, S. (1974) *J. Biochem. (Tokyo)* **76**, 935–947.
9. Vigne, R. & Jordan, B. R. (1977) *J. Mol. Evol.* **10**, 77–86.
10. Woese, C. R., Pribula, C. D., Fox, G. E. & Zablen, L. (1976) *J. Mol. Evol.* **7**, 197–213.
11. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
12. Stahl, D. A., Walker, T. A., Meyhack, B. & Pace, N. R. (1979) *Cell* **18**, 1133–1143.
13. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4**, 2527–2538.
14. Peattie, D. (1979) *Proc. Natl. Acad. Sci. USA* **74**, 4537–4541.
15. Donis-Keller, H. (1980) *Nucleic Acids Res.* **8**, 3133–3142.
16. Jordan, B. R., Galling, G. & Jourdan, R. (1974) *J. Mol. Biol.* **87**, 205–225.
17. Jordan, B. R., Galling, G. & Jourdan, R. (1973) *FEBS Lett.* **37**, 333–334.
18. MacKay, R. M., Spencer, D. F., Doolittle, W. F. & Gary, M. W. (1980) *Eur. J. Biochem.*, in press.
19. Luehrsen, K. R., Fox, G. E. & Woese, C. R. (1980) *Curr. Microbiol.* **4**, 123–126.
20. Nishikawa, U. & Takemura, S. (1978) *J. Biochem.* **84**, 259–266.
21. Erdmann, V. A. (1980) *Nucleic Acids Res.* **8**, r31–r47.
22. Tschudi, C. & Pirrotta, V. (1980) *Nucleic Acids Res.* **8**, 441–451.
23. Lu, A.-L., Steege, D. A. & Stafford, D. W. (1980) *Nucleic Acids Res.* **8**, 1839–1853.
24. Garrett, R. A., Douthwaite, S. & Noller, H. F. (1981) *Trends Biochem. Sci.*, in press.
25. MacKay, R. M., Gary, M. W. & Doolittle, W. F. (1981) *Nucleic Acids Res.* **8**, 4911–4917.
26. Hori, H., Osawa, S. & Iwabuchi, M. (1980) *Nucleic Acids Res.* **8**, 5535–5540.